UNIVERSITY OF TORONTO
3 1761 00073521 7

BF Spearman, Charles Edward
181 The proof and measurement of
573 association between two things

BF
181
573

RETURN TO
DEPARTMENT OF PSYCHOLOGY LIBRARY
UNIVERSITY OF TORONTO

Digitized by the Internet Archive
in 2007 with funding from
Microsoft Corporation

THE PROOF AND MEASUREMENT OF

ASSOCIATION

BETWEEN TWO THINGS.

By: C. Spearman,

from Amer Jour of Psy Vol. 15-1904

BF

181

S73

1001-01.100 page 1 of 100

612060

4.7.55

THE PROOF AND MEASUREMENT OF ASSOCIATION BETWEEN TWO THINGS

By C. Spearman

CONTENTS.

	Page
Introduction.	1
Part I. Elementary correlation, and "accidental deviations."	2
1. Requirements of a good method.	2
2. Standard methods explained.	6
3. Comparison by rank.	8
4. Auxiliary methods.	11
Part II. Correction of "systematic deviations."	17
1. Systematic deviations generally.	17
2. "Attenuation" by errors.	18
3. Limits of associative problems.	21
4. "Constriction" and "Dilation."	23
5. "Distortion."	24
6. Criticism of prevalent working methods.	25
7. Number of cases desirable for an experiment.	28

INTRODUCTORY.

All knowledge - beyond that of bare isolated occurrence - deals with uniformities. Of the latter, some few have a claim to be considered absolute, such as mathematical implications and mechanical laws. But the vast majority are only partial; medicine does not teach that smallpox is inevitably escaped by vaccination, but that it is so generally; biology has not shown that all animals require organic food, but that nearly all do so; in daily life, a dark sky is no proof that it will rain, but merely a warning; even in morality, the sole categorical imperative alleged by Kant was the sinfulness of telling a lie, and few thinkers since have admitted so much as this to be valid universally. In psychology, more perhaps than in any other science, it is hard to find absolutely inflexible coincidences; occasionally, indeed, there appear uniformities sufficiently regular to be practically treated as laws, but infinitely the greater part of the observations hitherto recorded concern only more or less pronounced tendencies of one event or attribute to accompany another.

Under these circumstances, one might well have expected that the evidential evaluation and precise mensuration of tendencies had long been the subject of exhaustive investigation

CONTENTS.

Page	
1	Introduction.
2	Part I. Elementary correlation, and 'accidental deviations'.
3	1. Requirements of a good method.
4	2. Standard methods explained.
5	3. Comparison by rank.
6	4. Auxiliary methods.
7	Part II. Correction of 'systematic deviations'.
8	1. Systematic deviations generally.
9	2. "Attenuation" by errors.
10	3. Limits of associative problems.
11	4. "Contraction" and "distortion."
12	5. "Distortion."
13	6. Criticism of prevalent working methods.
14	7. Number of cases desirable for an experiment.

INTRODUCTION.

All knowledge - beyond that of pure isolated occurrences - begins with uniformities. Of the latter, some few have a claim to be considered absolute, such as mathematical propositions and mechanical laws. But the vast majority are only probable; and medicine does not teach that amplex is invariably accompanied by vaccination, but that it is so generally; biology has not shown that all animals require organic food, but that nearly all do so; in daily life, a cork stopper is no proof that it will sink, but merely a warning; even in morality, the sociological theories advanced by Bentham were the richness of falling a lie, and few thinkers since have admitted so much as this to be valid universally. In psychology, more perhaps than in any other science, it is hard to find absolutely inflexible coincidences; occasionally, indeed, there appear uniformities sufficiently regular to be practically treated as laws, but strictly the greater part of the observations hitherto recorded concern only more or less pronounced tendencies of one event or attribute to accompany another.

Under these circumstances, one must not have expected that the evidential evaluation and precise measurement of tendencies has long been the subject of extensive investigation

and now formed one of the earliest sections in a beginner's psychological course. Instead, we find only a general naïve ignorance that there is anything about it requiring to be learnt. One after another, laborious series of experiments are executed and published with the purpose of demonstrating some connection between two events, wherein the otherwise learned psychologist reveals that his art of proving and measuring correspondence has not advanced beyond that of lay persons. The consequence has been that the significance of the experiments is not at all rightly understood, nor have any definite facts been elicited that may be either confirmed or refuted.

The present article is a commencement at attempting to remedy this deficiency of scientific correlation. With this view, it will be strictly confined to the needs of practical workers, and all theoretical mathematical demonstrations will be omitted; it may, however, be said that the relations stated have already received a large amount of empirical verification. Great thanks are due from me to Professor Haussdorff and to Dr. G. Lipps, each of whom have supplied a useful theorem in polynomial probability; the former has also very kindly given valuable advice concerning the proof of the important formulae for elimination of "systematic deviations."

At the same time, and for the same reason, the meaning and working of the various formulae have been explained sufficiently, it is hoped, to render them readily usable even by those whose knowledge of mathematics is elementary. The fundamental procedure is accompanied by simple imaginary examples, while the more advanced parts are illustrated by cases that have actually occurred in my personal experience. For more abundant and positive exemplification, the reader is requested to refer to the under cited research,¹ which is entirely built upon the principles and mathematical relations here laid down.

In conclusion, the general value of the methodics recommended is emphasized by a brief criticism of the best correlational work hitherto made public, and also the important question is discussed as to the number of "cases" required for an experimental series.

Part 1.

ELEMENTARY CORRELATION AND "ACCIDENTAL DEVIATION."

1. Requirements of a Good Method of Correlation.

(a) Quantitative expression.

The most fundamental requisite is to be able to measure our observed correspondence by a plain numerical symbol. There

1. "General Intelligence" determined and measured, to appear in a subsequent number of this Journal.

and now found one of the earliest decisions to a different
obolical course. Instead, we find only a single line
range that there is anything about it requiring to be tested. One
after another, laboratory series of experiments are executed as
published with the purpose of demonstrating some of the
between two events, therein the theories of psychology
reveals that the art of writing and reading correspondence
has not advanced beyond that of lay persons. The same
persons have been that the intelligence of the experiments is not
at all rightly understood, nor have any definite facts been
elicited that may be either confirmed or refuted.
The present article is a commentary of what is to be
remedy this deficiency of scientific correlation. In this view,
it will be strictly confined to the needs of medical officers, and
all theoretical mathematical demonstrations will be omitted;
it may, however, be said that the relations stated have already
received a large amount of scientific verification. Great thanks
are due to me to Professor Huxley and to Dr. G. H. Huxley,
each of whom have supplied a number of papers in psychology
probability; the former has also kindly given valuable
advice concerning the proof of the important theories for elimi-
nation of systematic deviations.
At the same time, and for the same reason, the meaning of
working of the various theories have been explained suffi-
ciently, it is hoped, to render them readily accessible even by
those whose knowledge of mathematics is elementary. The
fundamental procedure is so arranged that it is easily
examples, while the more advanced parts are illustrated by
cases that have actually occurred in my personal experience.
For the abundant and positive exemplification, the reader is
requested to refer to the under-aided research, which is entirely
built upon the principles and mathematical relations here laid
down.
In conclusion, the general value of the method of resem-
blance is emphasized by a critical analysis of the best avail-
able work hitherto made public, and also the important ques-
tion is discussed as to the number of cases required for an
experimental series.

Part I.

THEORY OF CORRELATION IN VARIOUS DEVIATIONS.

1. Determinants of the method of correlation.
The method of correlation is a statistical method.
The method of correlation is a statistical method.
The method of correlation is a statistical method.

1. General principles of correlation.
The method of correlation is a statistical method.

is no reason whatever to be satisfied either with vague generalities such as "large," "medium," "small," or, on the other hand, with complicated tables and compilations.

The first person to see the possibility of this immense advance seems to have been Galton, who, in 1886, writes: "the length of the arm is said to be correlated with that of the leg, because a person with a long arm has usually a long leg and conversely."¹ He then proceeds to devise the required symbol in such a way that it conveniently ranges from 1, for perfect correspondence, to 0 for entire independence, and on again to -1 for perfect correspondence inversely. By this means, correlations became comparable with other ones found either in different objects or by different observers; they were at last capable of leading to further conclusions, speculative and practical; in a word, they now assumed a scientific character.

Mathematically, it is clear that innumerable other systems of values are equally conceivable, similarly ranging from 1 to 0. One such, for instance, has been worked out and extensively used by myself (see pp 15 ff). It therefore becomes necessary to discuss their relative merits.

(b) The significance of the quantity.

Galton's particular system is defined and most advantageously distinguished from all the others by the important property, that if any number of arms, for instance, be collected which are all any amount, σ_a above the mean, then the corresponding legs will average $r\sigma_a$ above the mean (with a middle or "quartile" deviation (2) of $\sigma_l \sqrt{1-r^2}$); where σ_a = the quartile variation of the arms, σ_l = that of the legs, and r is the measure of the correlation.

But another - theoretically far more valuable - property may conceivable attach to one among the possible systems of values expressing the correlation; this is, that a measure might be afforded of the hidden underlying cause of the variations. Suppose, for example, that A and B both derive their money from variable dividends and each gets $1/x$ th. of his total from some source common to both of them. Then evidently their respective incomes will have a certain tendency to rise and fall simultaneously; this correspondence will in any of the possible systems of values always be some function $1/x$, but in only one of them will it actually be itself $1/x$; in such a favored case, if A and B get, say, 20% of their respective incomes from the common source, the correlation between these two incomes will also show itself as 0.20; and conversely, if A's income happens to be found correlated with that of B by 0.20, then

1. "Proceedings Royal Society of London", Vols XL and XLV.

2. Commonly, but misleadingly, termed the "probable error."

is no reason whatever to be satisfied either with vague generalities such as "large," "medium," "small," or, on the other hand, with complicated tables and calculations.

The first person to see the possibility of this immense advance seems to have been Galton, who, in 1888, writes: "the length of the arm is said to be correlated with that of the leg, because a person with a long arm has usually a long leg and conversely." He then proceeds to devise the "regression symbol" in such a way that it conveniently ranges from 1, for perfect correspondence, to 0 for entire independence, and on again to -1 for perfect correspondence inversely. By this means, correlations became comparable with other ones found either in different objects or by different observers; they were at least amenable of leading to further conclusions, speculative and practical; in a word, they assumed a scientific character. Mathematically, it is clear that innumerable other systems of values are equally conceivable, similarly ranging from 1 to 0. One such, for instance, has been worked out and extensively used by myself (see p. 17). It becomes necessary to discuss their relative merits.

(b) The significance of the quantity Galton's "regression symbol" is defined and most advantageously distinguished from all the others by the important property, that if any number of terms, for instance, be collected which are all any amount, $\frac{1}{2}$, above the mean, then the corresponding legs will average $\frac{1}{2}$ above the mean (with a middle or "quartile" deviation) of $\sqrt{1 - r^2}$; where r is the measure of the correlation.

But another - theoretically far more valuable - property may conceivably attach to one or more of the possible systems of values expressing the correlation; this is, that a measure might be afforded of the older underlying cause of the variations. Suppose, for example, that A and B both derive their money from some variable dividends and each gets $1/x$ th. of the total from some source common to both of them. Then evidently their respective incomes will have a certain tendency to rise and fall simultaneously; this correspondence will be any of the possible systems of values always as some function $1/x$, but in only one of them will it actually be $1/x$; in such a favored case, if A and B get, say, 5% of their respective incomes from the common source, the correlation between these two incomes will also show itself as 0.50; and conversely, if A's income happens to be fairly correlated with that of B by 0.30, then

I. "Proceedings Royal Society of London", Vols III and IV.

S. Cornhill, but mischievously, term it the "probable error."

there is a likelihood that 0.20 of A's income coincides with 0.20 of B, leaving to either 0.80 disposable independently. The observed correlation thus becomes the direct expression of the relative amount of underlying influences tending for and against the correspondence.

In the above imagined instance, this desirable expressiveness belongs to the same above system of values proposed by Galton (and elaborated by Pearson). But this instance is exceptional and fundamentally different from the normal type. Evidently, A and B need not necessarily derive exactly the same proportion of their incomes from the common source; A might get his 0.20 while B got some totally different share; in which case, it will be found that the correlation is always the geometrical mean between the two shares. Let B be induced to put all his income into the common fund, then A need only put in $0.20 \div 0.04 = 0.04$, to maintain the same correlation as before; since the geometrical mean between 0.04 and 1 is equal to 0.20.

Now, in psychological, as in most other actual correspondences, A and B are not to be regarded as in the fixed bisection of our first case, but rather as in the labile inter-accommodation of our second case. Hence A, in order to be correlated with B by $1/x$, must be considered to have only devoted $1/x^2$ (instead of $1/x$) of his arrangement to this purpose and therefore still have for further arrangement $1 - 1/x^2$, which will enable an independent correlation to arise of $\sqrt{1 - 1/x^2}$. In short, not Galton's measure of correlation, but the square thereof, indicates the relative influence of the factors in A tending towards any observed correspondence as compared with the remaining components of A tending in other directions.

(c) Accuracy.

From this plurality of possible systems of values for the measure of the correlation must be carefully distinguished the variety of ways of calculating any one of them. These latter again, have various advantages and disadvantages, of which the principal is their respective degrees of liability to "accidental deviation."

For, though the correlation between two series of data is an absolute mathematical fact, yet its whole real value lies in our being able to assume a likelihood of further cases taking a similar direction; we want to consider our results as a truly representative sample. Any one at all accustomed to original investigation must be aware how frequently phenomena will group themselves in such a manner as to convincingly suggest the existence of some law - when still more prolonged experiment reveals that the observed uniformity was due to pure hazard and has no tendency whatever to further repeat itself.

Luckily, this one great source of fallacy can be adequately

there is a likelihood that 0.20 of it is associated with 0.20 of A, leaving 0.40 which is independent. The observed correlation thus becomes the linear relation of the relative amount of underlying influence tending for and against the correspondence.

In the above example, the correlation coefficient expresses the degree to which the observed values are associated by chance. It is calculated by the formula: $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$. This formula is exact for the normal type of variation and fundamentally different from the normal type. In general, A and B need not necessarily have exactly the same proportion of their values from the common source; A might get his 0.20 while B got some relatively different share; in which case, it will be found that the correlation is always the geometric mean between the two shares. If B be thought to put all his income into the common fund, then a need only put in 0.20 of his income, so that in the same correlation as before, since the geometric mean between 0.20 and 1 is equal to 0.39, the correlation will be 0.39. In such cases, the correlation is not psychologically, but mathematically, a correspondence. A and B are not to be regarded as in the fixed relation of our first case, but rather as in the labile inter-association of our second case. Hence, in order to be correlated with B, A must be correlated to have only devoted 1/2 (instead of 1/3) of his expenditure to this purpose as there-fore will leave for further expenditure 1/2. This will enable us to interpret correlation as a ratio of 1/2 to 1/3. In short, no matter's measure of correlation, but the square thereof, indicates the relative influence of the factors in a tendency towards any observed correspondence as compared with the remaining components of tendency in other directions.

(b) Assumed.

From the principle of non-identity of values for the measure of the correlation must be stated; distinguished the variety of ways of calculating any one of them. These latter again, have various advantages and disadvantages, of which the principal is their respective degrees of liability to 'local' deviation.

For, though the correlation between two series of data is an absolute mathematical fact, yet its value may vary in our being able to measure a likelihood of further cases taking a similar direction; we must to consider our series as a truly representative sample. Any one of all specimens to original investigation would be a good but frequently phenomenon of a group themselves in some manner as a whole, but the existence of some law - when all I have said is an experiment reveals the observed tendency as due to the actual and has no reference to the further process itself. Finally, this one great source of error is adequately

eliminated, owing to the fact that such accidental deviations are different in every individual case (hence are often called the "variable errors") and occur quite impartially in every direction according to the known laws of probability. The consequence is that they eventually more or less completely compensate one another, and thus finally present an approximately true result. Such elimination, however, must always remain theoretically incomplete, since no amount of chance coincidence is absolutely impossible; but beyond certain limits it becomes so extremely unlikely that for practical purposes we can afford to neglect it. When a person loses 14 times running at pitch-and-toss, he can reckon that such a series would not occur by mere accident one in 9,999 times, and consequently he will feel justified in attributing the coincidence to some constant disturbing influence. Similarly, to estimate the evidential value of any other observed uniformity, we only require to know how nearly the odds against chance coincidence have approached to some such standard maximum as 9,999 to 1. But, as any standard must always be more or less arbitrary—some thinking it too lenient and others unnecessarily severe—it is usual to employ a formula giving not the maximum but the middle deviation or "probable error".¹ We may then easily find the probability of mere hazard from the following comparative table:

If the observed correlation
divided by the probable
error be.

	1	2	3	4	5	6
then the frequency of occurrence by mere hazard	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{23}$	$\frac{1}{143}$	$\frac{1}{1250}$	$\frac{1}{19000}$

Now, the smallness of this probable error depends principally upon the number of cases observed, but also largely upon the mathematical method of correlation. Though a faultiness in the latter respect can theoretically be made good by increasing the range of the observations, yet such increase is not always possible, and, besides, has other grave disadvantages which will be discussed later on. Other things being equal, therefore, the best method is that one which gives the least probable error. For the benefit of the reader, this probable error should always be plainly stated; nothing more is required than a rough approximation; for while it is highly important to distinguish between a deduction worth, say, 0.9999 of perfect certainty and one worth only 0.75, it would be a mere splitting of straws to care whether a particular experiment works out to a validity of 0.84 or to one of 0.85.

(d) Ease of application.

1. In the proper use of this expression.

eliminated, owing to the fact that such coincidental deviations are different in every individual case (force are often called the "variable error") and occur with equal probability in every direction according to the known laws of probability. The consequence is that they eventually cancel out or less completely cancel one another, and thus finally present an approximately zero result. Such elimination, however, is not always certain. It is entirely possible, since no amount of time or coincidence is absolutely impossible; but beyond certain limits it becomes an extremely unlikely event. Practical purposes we can afford to neglect it. Then a person takes 10 times 1000 mg at 1000 mg a day, he can expect that each time he will not only be more accurate one in 1000 times, and consequently he will feel justified in attributing the coincidence to some constant disturbing influence. Similarly, to estimate the evidential value of any other observed deviation, we can give a guide to know how rarely the same amount of coincidence have approached to some such statistical maximum as 1000 mg. But, as any statistic must always be more or less arbitrary, some thinking is too hasty and often unnecessarily covers it is usual to employ a formula giving not the maximum but the middle deviation or "probable error". It is then easily found the probability of error found from the following comparative table:

If the observed deviation is divided by the probable error we have:
 1.0 = 10%
 1.5 = 23%
 2.0 = 39%
 2.5 = 54%
 3.0 = 69%
 3.5 = 81%
 4.0 = 90%
 4.5 = 96%
 5.0 = 99%
 5.5 = 99.7%
 6.0 = 99.9%
 6.5 = 99.97%
 7.0 = 99.99%
 7.5 = 99.997%
 8.0 = 99.999%
 8.5 = 99.9997%
 9.0 = 99.9999%
 9.5 = 99.99997%
 10.0 = 99.99999%

Now, the smallness of this probable error becomes particularly upon the number of cases observed, but also largely upon the mathematical method of correlation. The method of correlation the latter method is theoretically better, but it is more difficult to apply, and, besides, has other grave disadvantages which will be discussed later on. Other things being equal, therefore, the best method is that one which gives the least probable error for the benefit of the reader, this probable error should always be plainly stated; nothing more is required to be known up-proximation; for while it is highly important to distinguish between a deviation worth, say, 0.001 of percent certainty and one worth only 0.01, it would be a mere trifling of error to give whether a test or other experiment is valid or of 0.84 or 0.85 of correlation.

I. in the present use of this expression.

The most accurate ways of calculation are generally somewhat difficult and slow to apply; often, too there occur circumstances under which they cannot be used at all. Hence in addition to a standard method, which must be used for finally establishing the principal results, there is urgent need, also of auxiliary methods capable of being employed under the most varied conditions and with the utmost facility.

But here a word of warning appears not out of place. For such auxiliary methods are very numerous and their results, owing to accidnets, will diverge to some extent from one another; so that the unwary, "self-suggested" experimenter may often be led unconsciously - but none the less unfairly - to pick out the one most favorable for his particular point, and thereby confer upon his work an unequivocality to which it is by no means entitled. Any departure from the recognized standard methods are only legitimate, either when absolutely necessary, or for mere preliminary work, or for indicating comparatively unimportant relations.

2. Standard Methods Explained.

(a) Correlations between variables that can be measured quantitatively.

This may be regarded as the normal type of correlation. Its standard method of calculation is that₂ discovered by Bravais,¹ in 1846, and shown by Pearson in 1896,² to be the best possible. Pearson terms this method that of "Product moments."

The formula appears most conveniently expressed as follows:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}}$$

where x and y are the deviations of any pair of characteristics from their respective medians,

xy is the product of the above two values for any single individual,

$\sum xy$ is the sum of such products for all the individuals,

$\sum x^2$ is the sum of the squares of all the various values of x,

$\sum y^2$ is similarly for y

and r is the required correlation.

A simple example may make this method clearer. Suppose that it was desired to correlate acuteness of sight with that of hearing, and that for this purpose five persons were tested as to the greatest distance at which they could read and hear a standard alphabet and sound respectively. Suppose the results to be;

1. "Memoires par divers savants" T, IX, Paris, pp255-332

2. "Phil. Trans., R.S., London" Vol.CLXXXVII, A, p.164.

Person	Sight	Hearing
A	6ft.	6ft.
B	7	11
C	9 (median)	12 (median)
D	11	10 (median)
E	14	8

	x	y	xy	x ²	y ²
then, we get					
A	- 3	- 4	+ 12	9	16
B	- 2	+ 1	- 2	4	1
C	0	+ 2	0	0	4
D	+ 2	0	0	4	0
E	+ 5	- 2	- 10	25	4

$$S_{xy} = \begin{pmatrix} +12 \\ -12 \end{pmatrix} \quad S_x^2 = 42 \quad S_y^2 = 25$$

$$+ 12 - 12$$

so that $r = \frac{+12 - 12}{\sqrt{42 \times 25}} = 0$, and there, thus, is no correspondence, direct or inverse.

The "probable error" between any obtained correlation and the really existing correspondence has been determined by Pearson, as being "with sufficient accuracy" when a fairly large number of cases have been taken,

$$= 0.674506 \sqrt{\frac{1 - r^2}{n(1 + r^2)}}$$

For discussion of correlation between characteristics whose distribution differs considerably from the normal probability curve as regards either "range" or "shewness," reference may be made to the works below.¹ It may be remarked that the method of "product moments" is valid, whether or not the distribution follow the normal law of frequency, so long as the "regression" is linear.

(b) Correlation between characteristics that can not be measured quantitatively.

In the example quoted by Galton, of correspondence between the length of arm and that of leg, it may be noted that the correspondence is proportional quantitatively; a long arm has a tendency to be accompanied by a leg not only long, but long to the same degree. Now, in many cases, such proportionality is by the nature of things excluded; a printed work is possibly remembered better than one heard; but, nevertheless, we cannot in accordance with the preceding formula, ascertain whether degrees of visuality are correlated to retentiveness of memory, seeing that in the former case there do not exist any degrees, a word being simply either seen or not seen. Perhaps even

1. Uany Yale: "Proc. R.S. London," Vol. LX, p477

Pearson: "Phil. Trans. R.S. London." Vol. CLXXXV, 1A, p71; Vol. CLXXXVI, 1 A, p343, and Vol. CXCI A, p.229.

G. Lipps: "Die Theorie der Collectivgegenstände," Wundt's Phil. Stud., Vol. XVII.

A
 B
 C
 D
 E

then we get

$$G = \frac{S_0}{\sqrt{\lambda}} \left(1 - \frac{S_0}{\sqrt{\lambda}} \right) = 0.728$$

Large number of cases have been taken,
Pearson, as being "with no definite answer" when a third
the really existing correspondence of the two series is by
The "probable error" between the obtained correlation and
✓ 42.4 25
so that $r = \frac{+18 - 18}{\dots} = 0$. and therefore, there is no cor-
respondence, direct or inverse.

$$\frac{100 - 1}{100 + 11.2} \times 60078.0 =$$

For discussion of correlation between characteristics and distribution differences, see the normal probability curve as measured either "by" or "between" the two. It may be made to the point where the method of "normal moments" is valid, whether or not the distribution is normal or frequency, so long as the "regression" is linear.

(b) Correlation between characteristics that can not be measured quantitatively.

In the example noted in (a), if correspondence between the length of arm and that of leg, it may be noted that the corresponding is essentially identical; a long arm has a tendency to be accompanied by a leg of only long, but long to the leg. In many cases, such correspondence is by the nature of things involved; a printed word is not remembered better than a heard; but, nevertheless, we cannot in accordance with a specific formula, ascertain whether degrees of similarity are involved to relative degrees of memory, seeing that the former is a form of non-abstract degree, a word being abstract and seen as a word, when even

PHIL. STU...
G. Dines:
Vol. CXXV
Lesson:
I. Ugarit

more numerous are those cases where proportionality does indeed exist, but practically will not admit of being measured; for instance, it is probable that conscientiousness is to some extent a hereditary quality, yet we cannot well directly determine whether brothers tend to possess precisely the same amount of it, owing to the fact that we cannot exactly measure it.

In all such cases we must confine ourselves to counting the frequencies of coexistence. We can easily find out how often seen and spoken words are respectively remembered and forgotten. It has proved quite feasible to divide the children of a school generally into "conscientious" and "non-conscientious," and then to measure how much brothers tend to be in the same division; When we have proved this simple association, we may provisionally assume correlation of quantity also; that is to say, if the "conscientious," generally speaking, have a particular degree of tendency to possess brothers likewise "conscientious," then boys with excessively tender scruples will have the same degree of tendency to possess brothers with similarly excessive tenderness, while those with only a moderate amount of virtue will be thus correlated with brothers also of only moderate virtue; further, the ethical resemblance may be expected to repeat itself in cousins, etc. only reduced in proportion as the kinship is diminished.

For measurement of this non-proportional association, a standard method, which may be termed that of "cross multiples," has been elaborated by Sheppard¹, Bramley-Moore, Filon, Lee, and Pearson. The formula is, unfortunately, too long and complicated to be usefully quoted in this place. It will be found in the under cited work² together with its probable error as determined by Pearson.³ In practice, it will generally have to be replaced by one of the more convenient methods to be next described.

3. Comparison by Rank.

This method of "cross multiples" is not only difficult and tedious of application, but also it gives a probable error nearly double that of "Product moments."

Now, it can often be altogether escaped in the case of quantities not admitted absolute measurement, by substituting instead comparison. This other way will be discussed at some length, as it has been largely used by myself and is believed chiefly responsible for some successful experiments. All characteristics may be collated from two quite distinct aspects: either (as in example of visual and auditory acuteness) by actual

1. "Phil. Trans." Vol. CXCII, A, p.141.

2. "Phil. Trans." Vol. CXCIV A, pp.2-7

3. Phil. Trans." Vol. CXCIV A, 10-14.

There is a certain amount of freedom of choice in the selection of the material to be used in the construction of the sentence. It is possible to select the material in such a way as to make the sentence as clear as possible. It is also possible to select the material in such a way as to make the sentence as obscure as possible. The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence.

The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence. It is possible to select the material in such a way as to make the sentence as clear as possible. It is also possible to select the material in such a way as to make the sentence as obscure as possible. The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence. It is possible to select the material in such a way as to make the sentence as clear as possible. It is also possible to select the material in such a way as to make the sentence as obscure as possible. The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence.

The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence. It is possible to select the material in such a way as to make the sentence as clear as possible. It is also possible to select the material in such a way as to make the sentence as obscure as possible. The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence. It is possible to select the material in such a way as to make the sentence as clear as possible. It is also possible to select the material in such a way as to make the sentence as obscure as possible. The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence.

8. Comparison of the two methods.

This method of choice is based on the principle of the least squares. It is possible to select the material in such a way as to make the sentence as clear as possible. It is also possible to select the material in such a way as to make the sentence as obscure as possible. The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence. It is possible to select the material in such a way as to make the sentence as clear as possible. It is also possible to select the material in such a way as to make the sentence as obscure as possible. The amount of freedom of choice is determined by the nature of the material and the purpose of the sentence.

mensuration, or else by order of merit: we might say that a student, A, obtained 8,000 marks in an examination, while B only got 6,000; or instead, we might say that A was third out of 100 candidates, while B was only 20th. Precisely the same method of calculation may be again used in the latter case, simply substituting the inverse ranks, 97, 80, etc., for the performances, 8,000, 6,000, etc.

(a) Disadvantages of the "Rank" method.

In the first place, it may be objected that the observed correlation would then only hold good for persons of the same average differences from one another. For assuming, say, acute sight to be correlated with acute hearing; then the order of merit of A, B and C, as regards sight, is more likely to remain unaltered as regards hearing also, when the difference in their respective powers of vision is extremely marked, than when they are practically equal on the latter head. But the more numerous the persons experimented on, the less will be the average difference of faculty; it might, therefore, be supposed that the correlation would become continually less perfect as the experiments were made more extensive. This, however, would be a fallacy: 100 experimental subjects compared together by "Rank" would on the whole actually show appreciably the same average correlation as 1,000, provided, that in either case the subjects are selected by chance; the amount of the correlation is not really dependent upon the difference between the grades, but upon the relation of this difference to the mean deviation; and both of these increase together with the number of subjects; On the other hand, the correlation will undoubtedly diminish if the subjects be all chosen from a more homogeneous class; in a select training school for teachers, for example, general intelligence will throughout show smaller correlation with other qualities, than would be the case in a college for quite average young men of the same age; but this fact applies just as much to comparison by "Measurement."

The next possible objection is that comparison by rank bases itself upon an assumption that all the subjects differ from one another by the same amount, whereas A may differ from B five times as much as B differs from C. But such an assumption would only take place, if correspondence by rank were considered to be wholly equivalent to that by measurement; no such assumption is made; the two aspects are recognized to be theoretically distinct, but advantage is taken of the fact that they give correlational values sensibly equivalent in amount. Even against the small existing discrepancy may be set off a deviation of the same order of magnitude which is incurred when using measurement itself, owing to the practical necessity of throwing the cases into a number of groups.

When the correlation is not significant, the null hypothesis is not rejected. If the correlation is significant, the null hypothesis is rejected and the alternative hypothesis is accepted. The test statistic is calculated as follows:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

where r is the correlation coefficient, X_i and Y_i are the individual scores on the two variables, $\bar{X}$ and $\bar{Y}$ are the means of the two variables, and n is the number of pairs of scores.

The test statistic is compared to the critical value from the t-distribution table for a given level of significance (e.g., 0.05) and degrees of freedom ($n-2$). If the test statistic is greater than the critical value, the null hypothesis is rejected and the alternative hypothesis is accepted. If the test statistic is less than the critical value, the null hypothesis is not rejected.

For example, suppose we have the following data:

X	Y
1	2
2	3
3	4
4	5
5	6

The mean of X is 3 and the mean of Y is 4. The standard deviation of X is 1.41 and the standard deviation of Y is 1.41. The correlation coefficient is calculated as follows:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = \frac{(1-3)(2-4) + (2-3)(3-4) + (3-3)(4-4) + (4-3)(5-4) + (5-3)(6-4)}{\sqrt{10 \cdot 10}} = \frac{10}{10} = 1.0$$

The test statistic is 1.0, which is greater than the critical value of 0.978 (for $n=5$ and $\alpha=0.05$). Therefore, the null hypothesis is rejected and the alternative hypothesis is accepted.

The third and only solid objection is that rank affords a theoretically somewhat less full criterion of correspondence than does measurement; and the force, even of this argument, disappears on considering that the two methods give appreciably the same correlational values.

(b) Advantages of the "Rank" method.

The chief of these is the large reduction of the "accidental error." In normal frequency curve, the outlying exceptional cases are much more spaced apart than are those nearer to the average; hence, any accident disturbing the position of these exceptional cases will have unduly great effect on the general result of the correlation; and owing to this inequality in the influence of the errors, the latter will not compensate one another with the same readiness as usual. Moreover, it is just these hyper-influential extreme cases where there is most likelihood of accidental errors and where there very frequently prevails a law quite different from that governing the great bulk of the cases. As regards the quantity of this gain by using rank (abstracting from the last mentioned point, which cannot well be estimated in any general manner) there should be no difficulty in calculating it mathematically. From a considerable amount of empirical evidence, the probable error when using the method of "product moments" with rank appears to become less than two-thirds of that given by the same method with measurement, and therefore only about one-third of that given by the method of "cross multiples."

The next advantage is that rank eliminates any disparity between the two characteristics compared, as regards their general system of distribution; such a disparity is often not intrinsic or in any way relevant, but merely an effect of the particular manner of gaining the measurement. By means of rank, a series presenting the normal frequency curve can be compared on even terms with another series whose curve is entirely different. This cannot well be done when using measurements. (See p.7).

Rank has also the useful property of allowing any two series to be easily and fairly combined into a third composite one.

(c) Conclusion.

From the practical point of view, it is so urgently desirable to obtain the smallest probable error with a given number of subjects, that the method of rank must often have the preference even when we are dealing with two series of measurements properly comparable with one another.

Theoretically, rank is at any rate preferable to such a hybrid and unmeaning correlation as that between essential measurements on the one side and mere arbitrary classification on the other. As the latter occur in most psychological correlations,

The third and only satisfactory method of determining the theoretical error is by the use of the method of least squares. This method is based on the assumption that the errors are distributed normally and that the mean of the errors is zero. The method of least squares is a statistical method which is used to find the best fit of a set of data points. It is based on the principle of minimizing the sum of the squares of the residuals. The residuals are the differences between the observed values and the values predicted by the model. The method of least squares is a powerful tool for analyzing data and for determining the best fit of a model to the data.

(b) Advantages of the "Least Squares" Method

The method of least squares has several advantages. First, it is a simple and easy-to-use method. Second, it is a statistical method which is based on the assumption that the errors are distributed normally. This assumption is often reasonable in practice. Third, the method of least squares is a powerful tool for analyzing data and for determining the best fit of a model to the data. Fourth, the method of least squares is a method which is used in many fields of science and engineering. Fifth, the method of least squares is a method which is used in many fields of business and economics. Sixth, the method of least squares is a method which is used in many fields of social science. Seventh, the method of least squares is a method which is used in many fields of medicine and biology. Eighth, the method of least squares is a method which is used in many fields of agriculture and forestry. Ninth, the method of least squares is a method which is used in many fields of engineering and technology. Tenth, the method of least squares is a method which is used in many fields of art and design. Eleventh, the method of least squares is a method which is used in many fields of music and drama. Twelfth, the method of least squares is a method which is used in many fields of sports and recreation. Thirteenth, the method of least squares is a method which is used in many fields of education and training. Fourteenth, the method of least squares is a method which is used in many fields of government and public administration. Fifteenth, the method of least squares is a method which is used in many fields of law and justice. Sixteenth, the method of least squares is a method which is used in many fields of religion and spirituality. Seventeenth, the method of least squares is a method which is used in many fields of philosophy and ethics. Eighteenth, the method of least squares is a method which is used in many fields of psychology and behavior. Nineteenth, the method of least squares is a method which is used in many fields of sociology and anthropology. Twentieth, the method of least squares is a method which is used in many fields of history and culture.

The next advantage of the method of least squares is that it is a method which is used in many fields of science and engineering. The method of least squares is a powerful tool for analyzing data and for determining the best fit of a model to the data. It is based on the principle of minimizing the sum of the squares of the residuals. The residuals are the differences between the observed values and the values predicted by the model. The method of least squares is a method which is used in many fields of science and engineering. It is a method which is used in many fields of business and economics. It is a method which is used in many fields of social science. It is a method which is used in many fields of medicine and biology. It is a method which is used in many fields of agriculture and forestry. It is a method which is used in many fields of engineering and technology. It is a method which is used in many fields of art and design. It is a method which is used in many fields of music and drama. It is a method which is used in many fields of sports and recreation. It is a method which is used in many fields of education and training. It is a method which is used in many fields of government and public administration. It is a method which is used in many fields of law and justice. It is a method which is used in many fields of religion and spirituality. It is a method which is used in many fields of philosophy and ethics. It is a method which is used in many fields of psychology and behavior. It is a method which is used in many fields of sociology and anthropology. It is a method which is used in many fields of history and culture.

The next advantage of the method of least squares is that it is a method which is used in many fields of science and engineering. The method of least squares is a powerful tool for analyzing data and for determining the best fit of a model to the data. It is based on the principle of minimizing the sum of the squares of the residuals. The residuals are the differences between the observed values and the values predicted by the model. The method of least squares is a method which is used in many fields of science and engineering. It is a method which is used in many fields of business and economics. It is a method which is used in many fields of social science. It is a method which is used in many fields of medicine and biology. It is a method which is used in many fields of agriculture and forestry. It is a method which is used in many fields of engineering and technology. It is a method which is used in many fields of art and design. It is a method which is used in many fields of music and drama. It is a method which is used in many fields of sports and recreation. It is a method which is used in many fields of education and training. It is a method which is used in many fields of government and public administration. It is a method which is used in many fields of law and justice. It is a method which is used in many fields of religion and spirituality. It is a method which is used in many fields of philosophy and ethics. It is a method which is used in many fields of psychology and behavior. It is a method which is used in many fields of sociology and anthropology. It is a method which is used in many fields of history and culture.

(c) Limitations of the "Least Squares" Method

The method of least squares has several limitations. First, it is a statistical method which is based on the assumption that the errors are distributed normally. This assumption is often reasonable in practice, but it is not always true. Second, the method of least squares is a method which is used in many fields of science and engineering. It is a method which is used in many fields of business and economics. It is a method which is used in many fields of social science. It is a method which is used in many fields of medicine and biology. It is a method which is used in many fields of agriculture and forestry. It is a method which is used in many fields of engineering and technology. It is a method which is used in many fields of art and design. It is a method which is used in many fields of music and drama. It is a method which is used in many fields of sports and recreation. It is a method which is used in many fields of education and training. It is a method which is used in many fields of government and public administration. It is a method which is used in many fields of law and justice. It is a method which is used in many fields of religion and spirituality. It is a method which is used in many fields of philosophy and ethics. It is a method which is used in many fields of psychology and behavior. It is a method which is used in many fields of sociology and anthropology. It is a method which is used in many fields of history and culture.

the only other resource would be to avoid measurements altogether by using the method of "cross multiple." But this trebles the size of the probable error, and therefore renders it necessary that the subjects should be no less than nine, times as numerous; such an enormous increase, even if possible, would generally be accompanied by disadvantages infinitely outweighing the supposed theoretical superiority of method.

The above advantages are still further enhanced whenever dealing with one-sided frequency curves, such as are furnished by most mental tests. For in these cases the great bulk of influence upon the resulting correlation is derived exclusively from the very worst performances and is consequently of a specially doubtful validity.

In short, correlation by rank, in most cases a desirable procedure, is for short series quite indispensable, rendering them of equal evidential value to much longer ones treated by other ways. Luckily, it is precisely in short series that gradation by rank is practically attainable.

(4) Auxiliary Methods.

These, as has been said, are only for use when there is adequate reason for not employing the above "standard" methods. Any number are devisable. Their resulting correlational values do not quite coincide with those found by the standard says, but nearly enough so for most practical purposes.

(a) Auxiliary methods of Pearson

Several very ingenious and convenient ones are furnished by him,¹ but all of similar type and requiring the same data as that of "cross-multiple."² They are therefore for use when the compared events do not admit of direct quantitative correlation. The following appears to combine facility and precision to the greatest degree:

$$r = \sin \frac{\pi}{2} \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

where the two compared series of characteristics, say P and Q are each divided into two (preferably about equal) classes; if the case is one where quantity exists but cannot be absolutely measured, P II will comprise the instances in which P is in manifest deficiency; but if the compared characteristics essentially exclude quantity, P II becomes the instances where P is

1. "Phi. Trans. R.S.L.," Vol. CXCIV, A, pp 1 and 79.

2. They are all refinements of the original formula, $r = \frac{ad - bc}{ad + bc}$ published by Yule, Proc. R.S.L., Vol. LXVI, p.23.

the only other resource would be a very small number of
gather by using the method of random sampling. This method
troubles the size of the problem. It is necessary to
necessary that the subjects should be as well as possible, and
numerous; such an enormous number, even if possible, would
generally be accompanied by the risk of distortion of the
the proposed theoretical understanding of the problem.
The above advantages are still further added by the
dealing with one of the many subjects, and the method of
by post mental tests. For it is not possible to have a
influence upon the results of the tests, and the results
from the very worst method of testing, and the results
especially in the case of the tests.
In short, correlation of the results of the tests
cannot be for short. It is not possible to have a
of equal statistical value. It is not possible to have a
were. Finally, it is not possible to have a
by which the results of the tests are obtained.

1.1. Auxiliary Methods.

There are two main methods of testing. The first is the
method of random sampling. The second is the method of
any number of subjects. The results of the tests are
do not differ from the results of the tests. The results
but they are not the same as the results of the tests.

1.2. Auxiliary Methods of Testing.

Several very important methods of testing are the
method of random sampling. The second is the method of
any number of subjects. The results of the tests are
do not differ from the results of the tests. The results
but they are not the same as the results of the tests.

where the two components are the same. The results of the tests are
do not differ from the results of the tests. The results
but they are not the same as the results of the tests.

1. "Yield" Test.
2. "Yield" Test.
3. "Yield" Test.

absent; similarly Q. Then,

a	=	the number of times that P I is accompanied by Q I
b	"	" " " " P II " " Q II
c	"	" " " " P I " " Q II
d	"	" " " " P II " " Q II

If $a + b$ is not very unequal to $c + d$, the probable error may be taken at about $1.1/\sqrt{n}$, where n = the number of instances in the whole of P or of Q.¹

Returning to our previous illustration, suppose that it was desired positively to ascertain the merits of instruction by writing and by word of mouth respectively. Ten series, each consisting of ten printed words, have been successively shown to a class of twenty children, who each time had to write down by memory as many as they could. The experiment was next repeated, but reading the words aloud instead of showing them. Of the 2,000 visual impressions 900 were correctly remembered, while of the same number of auditory ones only 700 were retained.

Call the visual impressions	P I
" " auditory	P II
" " remembered	Q I
" " forgotten	Q II

then $a = 900$, $b = 700$, $c = 1,100$, $d = 1,300$, and

$$r = \sin \frac{\pi}{2} \frac{\sqrt{900} \times \sqrt{1,300} + \sqrt{700} \times \sqrt{1,100}}{\sqrt{900} \sqrt{1,300} + \sqrt{700} \sqrt{1,100}} = 0.16$$

the probable error then comes to $1.1/\sqrt{4,000}$ = nearly 0.02, or about 1/8 of the above correlation; so that the latter would not occur by mere chance once in 100,00 times.

We thus see that there is at any rate good prima facie evidence of some superiority on the part of the visual sense. Also, if the experiment has been fairly executed and adequately described, any subsequent verification under sufficiently similar conditions, by other experimenters, should result in a concordant correlation, probably between 0.04 and 0.28.

Moreover, we have obtained a direct estimate of the importance of this apparent superiority of the visual sense; for the square of the correlation amounts to 0.025; so that of the various causes here tending to make the children remember some words better than others, the difference of sense impressed comes to about one fortieth part (see p. 4).

1. More accurately, $\sin 0.1686 \pi (1 - r^2) \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$

(b) Method of proportional changes.

This is very often convenient, being especially applicable to a large number of psychological experiments, and so easy that the result can be approximately seen on inspection. Here,

$$r = \frac{3}{2} \cdot \frac{a - b}{a + b}$$

where a = the number of cases that have changed in accordance with the supposed correspondence, and b = the number that have changed in contradiction of it. The probable error again comes to $\frac{1.1}{\sqrt{n}}$

Suppose, for example, we were demonstrating that intellectual fatigue may be satisfactorily investigated by the method of Griessbach.² With this view, we have applied his test to 100 boys before and after their lessons. In the latter case 68 of them have presented the expected duller sensitivity, but 32, on the contrary, have shown a finer discrimination than before work.

Now, clearly, had the correspondence been perfect, all the hundred would have become worse.³ Thus,

$$r = \frac{3}{2} \cdot \frac{68 - 32}{100} = 0.54.$$

As the probable error comes to 0.11, our imaginary correlation is five times greater, and therefore would not have occurred by mere accident more than once in 1,250 times; so that we become practically certain that the sensitivity of the skin really does measure fatigue.

It now becomes easy to compare the quantity of this fatigue at different stages of work. Let us say that further experiments, after lessons lasting one hour longer than before, showed the correlation had risen to 0.77. Thereby we see that the influence of fatigue swells from 0.54² to 0.77,² that is, from being 1/5 to being 3/5 of all the sources of variation in cutaneous sensitivity. Such a result has a very different scientific significance from, say, any conclusion that the average sensory threshold had enlarged by so many more millimetres.

1. Hence, when the correlation is very complete, say over 0.75 the above formula gives appreciably too large values; as the amount reaches 0.90 and 1, the first factor must be reduced from 3/2 to 5/4 and 1 respectively.

2. This, as is well known, consists in determining the least distance apart at which two points of contact can be distinguished as being double and not single.

3. Assuming that is to say, that all the boys become fatigued by

(b) Method of correlation

This is very often convenient, being a lot less difficult than a large number of experiments. The results can be compared with those of other experiments.

$$\frac{1}{2} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

where $\frac{1}{2}$ - the number of cases in which the results are the same, $\frac{1}{4}$ - the number of cases in which the results are different. The probability of the results being the same is $\frac{1}{2}$.

For example, we have a coin which is tossed 100 times. The results may be 50 heads and 50 tails. With this result we have a probability of $\frac{1}{2}$ of getting 50 heads and 50 tails. The probability of getting 50 heads and 50 tails is $\frac{1}{2}$. The probability of getting 50 heads and 50 tails is $\frac{1}{2}$.

Now, clearly, the results of the experiment are not the same. The results of the experiment are not the same.

$$\frac{1}{2} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

As the probability of error is $\frac{1}{4}$, the probability of error is $\frac{1}{4}$. The probability of error is $\frac{1}{4}$. The probability of error is $\frac{1}{4}$. The probability of error is $\frac{1}{4}$.

It now becomes easy to compare the results of the experiment at different stages of error. The results of the experiment at different stages of error are compared. The results of the experiment at different stages of error are compared. The results of the experiment at different stages of error are compared.

1. Hence, when the correlation is very low, the results of the experiment are compared. The results of the experiment are compared. The results of the experiment are compared.

2. This is well known, and the results of the experiment are compared. The results of the experiment are compared. The results of the experiment are compared.

3. Assuming, that is to say, that the results of the experiment are compared. The results of the experiment are compared. The results of the experiment are compared.

Moreover, our test can be easily and precisely compared with any of the various other recommended procedures, being more reliable than all which present smaller correlations and vice versa.

(c) Method of class averages.

It often happens that measurements (or ranks) are known but not in such a way as to be able to use either the method of "product moments" or even any of the methods of Pearson. Under such circumstances, I have found it very useful to be able to apply the following relation:

$$r = \frac{d}{D}$$

where d is the observed difference between the average measurement (or rank) of the P's accompanied by Q I and that of those accompanied by Q II, and D is the greatest difference that was possible (such as would have occurred, had the correspondence been perfect). If Q has been divided into two about equal portions, D will be equal to twice the middle or "quartile" deviation from the average in the whole series P ; while if Q has been divided after the usual fashion into three such portions, only the two outer ones can be used and then $D = 2.87$ times the above middle deviation (again taken in the whole series P)

Suppose, for example, that we wish to ascertain whether the well known test of "reaction-time" gives any indication as to the person's general speed of movement. We try a hundred persons both in reaction-time and in speed of running 50 yards. Then we divide the reaction-time records into two classes, I containing all the quickest performers and II all the slowest. We now see how long these two classes of reactors took respectively to run the fifty yards, and what was the middle deviation from the average among all the runners taken together. Let us put the average of class I at 6 seconds, that of class II at 6.5 seconds, and the general middle deviation at 1.1 seconds. Then

$$r = \frac{6.5 - 6}{2 \times 1.1} = 0.23$$

The evidential value of the result is given approximately, even for small values of n , by the following relation:

$$\text{Probable error} = \frac{1.17}{\sqrt{n}} \cdot \frac{\sqrt{n} \pm 1}{\sqrt{n} + 2}$$

where n is the total number of cases considered. In the three-

fold instead of twofold division, the probable error becomes nearly

$$\frac{1.4}{\sqrt{n}} \quad \frac{\sqrt{n} - 1}{\sqrt{n} - 2}$$

In the above instance, we find that the observed correlation is little over double the probable error; as so much would turn up about once in six times by mere accident, the evidence is not at all conclusive. Therefore we must either observe many more cases - 600 would be necessary to reduce the probable error to 1/5th of the correlation - or else we must find a better method of calculation. If rank had been employed instead of measurement, the evidence would already have been fairly good, and could have been put beyond all reproach by the addition of another 150 observations. If rank had been employed in conjunction with the method of "product moments" or that of "rank differences," the required smallness of probable error could have been obtained by as few as 36 cases in all!

The method of "class averages" is especially valuable in deciphering the results of other investigators, where the average performances and the middle deviations are usually given (in good work), but not the data required for any of the other methods.

(d) Method of rank differences.

This method appears to deserve mention also, seeing that it seems to unite the facility of the auxiliary methods with a maximum accuracy like that given by "product moments". It depends upon noting how much each individual's rank in the one faculty differs from his rank in the other one; evidently this will be nil when the correlation is perfect, and will increase as the correlation diminishes.¹

1. This general idea seems to have been first due to Binet and Henri ("La fatigue intellectuelle" p.252-261), who, however do not work it out far enough to obtain any definite measure of correlation. Accordingly, Binet makes little further attempt in later research (L'annee psychologique, Vol. IV) to render it of service, and soon appears to have altogether dropped it (L'annee psychologique, Vol. VI.).

The same idea occurred to myself and was developed as above, without being at the time acquainted with the previous work in this direction by Binet and Henri. In obtaining the above formulae I was greatly assisted by Dr. G. Lipps' showing generally that when an urn contains n balls numbered 1, 2, 3, ..., n , respectively; and when they are all drawn in turn (without being replaced); and when the difference is each time noted between the number on the ball and the order of its drawing; then the most probable (or middle) total sum of such differences, added together without regard to sign, will be

$$= \frac{n^2 - 1}{3}$$

Previously I had only calculated this value for each particular size of n required by myself. Prof. Hausdorff further showed, generally, that such sum of differences will present a mean square deviation (from the above most probable value)

$$= \sqrt{\frac{(n+1)(2n^2+7)}{45}}$$

The relation is as follows:

$$R = 1 - \frac{3Sd}{n^2 - 1}$$

where Sd is the sum of the differences of rank for all the individuals,

n is the total number of individuals,
and R is the required correlation.

The probable error will then be approximately, even for small values of n , $\approx 0.4/\sqrt{n}$.

To take again the example from p.9, we number the five persons according to their order of merit in hearing and seeing respectively.

person	seeing rank	hearing rank	difference
A	1	1	0
B	2	4	2
C	3	5	2
D	4	3	1
E	5	2	3

$$Sd = 8$$

so that

$$R = 1 - \frac{3 \cdot 8}{25 - 1} =$$

and again we find that there is no correlation, direct or inverse.

This method, though very accurate and pre-eminently quick in application, has unfortunately four serious disadvantages.

It can be only used for ranks, and not immediately for measurements.

The probable error given is only that showing how great correlations may be expected from pure accident when there is no really existing correspondence between the two characteristics. It does not (like Pearson's probable error for the method of "product moments") directly show how much the observed correlation may be expected to differ by accident from any correspondence that does exist.

The various possible values of Sd are found to fall into a frequency curve of marked asymmetry; so that we cannot (as in all the other methods here given) take the minus values of R as representing so much inverse correlation. This defect could be remedied mathematically; but there are also other respects in which this side of the frequency curve appears unsuitable for our purpose, so that it is better to treat every cor-

1. This formula becomes slightly incorrect, whenever two or more individuals are bracketed as having precisely the same rank; but the consequent error is usually to be worth considering.

The relation is as follows:

$$R = \frac{1}{n} \sum_{i=1}^n \frac{d_i}{d_i + 1}$$

where d_i is the sum of the differences of rank for all the individuals.
 n is the total number of individuals.
 and R is the required correlation.
 The probable error will then be approximately, even for small values of n , $\pm 0.67/\sqrt{n}$.
 To take again the example from p. 10, we number the five persons according to their order of merit in hearing, respectively.

person	hearing rank	difference
A	1	0
B	2	1
C	3	2
D	4	3
E	5	4

$$R = \frac{10}{25} = 0.4$$

so that

$$R = 1 - \frac{2}{25 - 1} = \frac{23}{24}$$

and again we find that there is no correlation, direct or inverse. This method, though very accurate and practically applicable in application, has unfortunately four serious disadvantages. It can be only used for ranks, and not for arbitrary measurements.

The probable error given is only that showing how great correlations may be expected from data significant when there is no really existing correspondence between the two measurements. It does not (like Pearson's probable error for the method of "product moments") directly show how much the observed correlation may be expected to differ by accident from any correspondence. What does it show?

The various possible values of R are found to fall into a frequency curve of marked asymmetry, and we cannot (as in all the other methods here given) take the three values of R as representative of an inverse correspondence. This defect could be remedied mathematically; but there are two other respects in which this side of the frequency curve appears unsuitable for our purpose. (a) It is better to treat every correlation as a random variable, and (b) it is better to treat every correlation as a random variable.

1. This formula becomes slightly incorrect, however, when two or more individuals are presented as having exactly the same rank, but the consequent error is usually to be worth considering.

relation as positive (which can always be done by, if necessary, inverting the order of one of the series).

Finally, this value R is not numerically equivalent to the " r " found by all the other methods, but for chance distributions appears $= \sqrt{r^2}$. So far, the proof of this relation is only empirical, but it rests on a large number of cases taken, however, only between 0.20 and 0.60. If it be accepted r can at once be found from the following table;

R	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1
r	0.13	0.22	0.34	0.44	0.54	0.63	0.71	0.79	0.86	0.93	1

Part II.

Correction of "Systematic Deviations."

1. Systematic Deviations Generally.

In the first part, we have seen that any correlational experiments however extensive, can only be regarded as a "sample" out of the immense reality, and will consequently present a certain amount of accidental deviation from the real general tendency; we have further seen that this accidental deviation is measurable by the "probable error" whose determination, therefore becomes an indispensable requisite to all serious research.

But now we are in danger of falling from Scylla into Charybdis. For after laboriously compiling sufficient cases and conscientiously determining the probable error, there exists a very human tendency to cease from labor and inwardly rejoice at having thus risen from common fallacious argument to the serene certainty of mathematics. But whether or not such complacency may be justifiable in pure statistical inquiry, it is at any rate altogether premature in the kind of research that we are at present contemplating; we are not dealing with statistics, but with a line of work so fundamentally different that it may be aptly distinguished by the term of "statisticoids." Here the accidental deviation is not the sole one, nor even the most momentous; there are many other enemies who are unmoved by the most formidable array of figures. These consist in such deviations as, instead of merely being balanced imperfectly, lie wholly on the one side or the other. As in ordinary measurements, so too in correlation, we may speak, not only of "accidental" "variable" or "compensating" inaccuracies, but also of "systematic," "constant," or "non-compensating" ones.

These systematic deviations are of very varied nature, the most insidious being as usual self suggestion. To take, for instance, one of our recent examples, suppose that we have applied the Griessbach test to a number of children before and

relation as positive (which can
inverting the order of one of the
- Finally, this is the
"r" found in all the other
tion appears -
applied to the
every, and
once for
R.O. 0.10 0.10 0.10
R.O. 0.10 0.10 0.10

Correlation of the
Part II.

1. Systemic is a
In the first
ment is
out of the
certain amount of
tenary; and
is
therefore
However,
dit. The
correlation
very
at
correlation
at
no
it
Here
most
moved
in
generally
and
of
But
There
not
instance
applied

after their lessons, and have found the desired correlation between fatigue and cutaneous insensitivity it still remains exceedingly difficult to convince ourselves that we executed our tests entirely without favor or affection; for it is almost impossible to determine a series of sensory thresholds without some general tendency, either to bring them towards the desired shape, or else - endeavoring to escape such bias - to force them in the opposite direction. To convince others of our impartiality may be harder still. Even this sort of deviation is to be remedied by our proposed exact method of procedure for by it we obtain perfectly definite results which any impartial experimenters may positively corroborate or refute.

2. "Attenuation" by Errors.

From page 3 it will be obvious that a correlation does not simply depend on the amount of concurring factors in the two compared series, but solely on the proportion between these concurring elements on the one hand and the discording ones on the other. In our example it did not matter whether A and B each had one pound or a thousand pounds in the common funds but only whether the amount was a small or large fraction of their whole incomes. If the discordance, $1-x$, be nil, then the concordance x is thereby perfect, that is, $= 1$; and if the influence of the discordant elements be sufficiently increased, then any concordance will eventually become infinitely small.

To consider a still more concrete example suppose three balls to be rolled along a well-kept lawn; then the various distances they go will be almost perfectly correlated to the various forces with which they were impelled. But let these balls be cast with the same inequalities of force down a rough mountain side; then the respective distances eventually attained will have but faint correspondence to the respective original momenta.

Thus it will be clear that here the accidental deviations have a new consequence simultaneous with, but quite distinct from, that discussed in the last chapter. For there, they impartially augmented and diminished the correlation, tending in a prolonged series to always more and more perfectly counterbalance one another; and in ordinary measurements, this is their sole result. But here in correlations, they also have this new effect which is always in the direction of "attenuating" the apparent correspondence and whose amount, depending solely on the size of the middle error, cannot be in the least eliminated by

I. This fact has already been mathematically expressed in the last chapter by the value of correlation between two series being proportional (inversely) to the value of the middle deviation inside the series (see p. 15).

after their lessons, and have found the desired correlation between fatigue and attention. It still remains exceedingly difficult to convince ourselves that we executed our tests entirely without favor or effect; for it is almost impossible to determine a series of sensory thresholds without some general tendency, either to bring them towards the desired range, or else - and more frequently - to bring them in the opposite direction. To convince others of our impartiality may be harder still. Even this sort of deviation is to be remedied by our proposed exact method of procedure for by it we get in perfectly definite results which any impartial experimenters may positively corroborate or refute.

2. "Attention" by errors. From page 2 it will be obvious that a correlation does not simply depend on the amount of concentrating power in the compared series, but solely on the proportion between the corresponding elements on the one hand and the corresponding ones on the other. In our example it did not matter whether and how much one turned on a thousand points in the comparison series, but only whether the amount was a small or large fraction of their whole incomes. If the discrepancy, x , be all, then the concordance x is a very small part, that is, $x/1$; and if the difference of the discordant elements be sufficiently increased, then any concordance will eventually become infinitesimally small.

To consider a still more concrete example, suppose three balls to be rolled along a half-foot lawn; then the various distances they go will be almost perfectly correlated to the various forces with which they were impelled. But let these balls be cast with the same inequalities of force down a rough lawn - then the respective distances are not even approximately correlated; they will have but faint correspondence to the respective original momenta.

Thus it will be clear that here the accidental deviations have a new consequence of influence with the other distant factor that discussed in the last chapter. In this case, what is really suggested and illustrated is the correlation, leading in a long series to always more and more perfectly uncorrelated one another; and in ordinary measurement, this is their sole result. But here in correlations, they may have this new effect which is always in the direction of "attention" and "preparation" and whose amount, depending solely on the correspondence and whose amount, depending solely on the size of the whole error, cannot be in the least diminished by

1. This has already been mentioned in the last chapter by the value of correlation between two series being proportional (inversely) to the value of the middle deviation (p. 15).

any prolongation of the series. The deviation has thus become general or "systematic."

Now suppose that we wish to ascertain the correspondence between a series of values, p , and another series, q . By practical observation we evidently do not obtain the true objective values, p and q , but only approximations which we will call p' and q' . Obviously, p' is less closely connected with q' than is p with q . For the first pair only correspond at all by the inter-mediation of the second pair; the real correspondence between p and q shortly r_{pq} , has been "attenuated" into $r_{p'q'}$.

To ascertain the amount of this attenuation and thereby discover the true correlation it appears necessary to make two or more independent series of observations of both p and q . Then

$$r_{pq} = \frac{r_{p'q'}}{\sqrt{r_{p'p'} \cdot r_{q'q'}}$$

where $r_{p'q'}$ = the mean of the correlations between each series of values obtained for p with each series obtained for q .

$r_{p'p'}$ = the average correlation between and another of these several independently obtained series of values for p .

$r_{q'q'}$ = the same as regards q .

and r_{pq} = the required real correlation between the true objective values of p and q .

Thus, if for each characteristic two such independent series of observations be made, say $p_1 p_2 q_1$ and q_2 then the true

$$r_{pq} = \frac{r_{p_1 q_1} + r_{p_1 q_2} + r_{p_2 q_1} + r_{p_2 q_2}}{4}$$

$$\sqrt{(r_{p_1 p_2} \times r_{q_1 q_2})}$$

Should circumstances happen to render say p_1 much more accurate than p_2 , then the correlations involving p_1 will be considerably greater than those involving p_2 . In such case, the numerator of the above fraction must be formed by the geometrical instead of by the arithmetical mean; here by the accidental errors of the respective observations cease to eliminate one another and therefore double their final influence; they also introduce an undue diminution of the fraction.

In some exceptional and principally very theoretical cases, it may happen that either of the actual measurements, say p_1 is

- 1 By an inversion of the above formula, the correlation between two series of observations will be found a useful measure of the accuracy of the observation.

connected with q' (or q) quite independently of p or any other link common to p' . Then, the correlation $r_{p'q'}$ will be to that extent increased without any proportional increase in $r_{p'p'}$;

hence our above formula will fallaciously present too large a value.

A greater practical difficulty is that of obtaining two series sufficiently independent of one another. For many errors are likely to repeat themselves; even two separate observers are generally, to some extent warped by the same influences; we are all imposed on by, not only the "Idola Specus," but also the "Idola Tribus" and the "Idola Fori." In such case, the above formula is still valid, only, its correction does not go quite far enough, - a fallacy at any rate on the right side.

An actual instance will best show the urgent necessity of correcting this attenuation. In a correlation between two events, say P and Q , I obtained three independent observations both of P and of Q . The average correlation for those of P with those for Q was 0.38 ($= r_{p'q'}$); the average correlation of those for P with one another was 0.58 ($= r_{p'p'}$); the same for Q was 0.22 ($= r_{q'q'}$). Therefore, the correspondence

between the real events, P and Q , comes by reckoning to

$$\frac{0.38}{\sqrt{0.58 \times 0.22}} \approx \text{approximately } 1; \text{ so that the correspondence}$$

instead of being merely 0.38, appeared to be absolute and complete.

Attenuation by errors can also be corrected in another manner, which has the great advantage of an independent empirical basis, and therefore of not being subject to either of the two above mentioned fallacies besetting the other method. Hence, when the results coincide both ways, the fallacies in question may thereby be considered as disproved, for it is very unlikely that they should both be present and in such proportions as to exactly cancel one another. In this method, instead of directly employing the values $p_1 p_2 p_3$, etc., we amalgamate them into a single list; by this means we clearly eliminate some portion of the individual observational errors, and thereby we cause any really existing correspondence to reveal itself in greater completeness. Now, this increase in correlation from this partial elimination of errors will furnish a measure of the increase to be expected from an entire elimination of errors. Assuming the mean error to be inversely proportional both to this increase in the correlation and to the square root of the number of lists amalgamated, the relation will be:

$$r_{pq} = \frac{\sqrt[4]{mn \cdot r_{p'q'} - r_{p'q'}}}{\sqrt[4]{mn - 1}}$$

connected with p' (or p) quite independently of p or any other link common to p' & p. Then, the correlation p'p will be to that extent increased without any proportional increase in p'p;

hence our above formula will fallaciously present too large a value. A greater practical difficulty is that of obtaining two series sufficiently independent of one another. For many reasons, we are likely to repeat ourselves; even two separate observers are generally, to some extent warped by the same influences; we are all imposed on by, not only the "Idola Speciei," but also the "Idola Tribus" and the "Idola Temporis." In such cases the above formula is still valid, only, its correction does not go quite far enough, - a fallacy at any rate on the right side.

An actual instance will best show the urgent necessity of correcting this attenuation. In a correlation between two events, say P and Q, I obtained three independent observations both of P and of Q. The average correlation for those of P with those for Q was 0.45 ($= r_{p'q}$); the average correlation of those for P with one another was 0.58 ($= r_{p'p}$); the same for Q was 0.82 ($= r_{q'q}$). Therefore, the correction

between the real events, P and Q, comes by reasoning to

$$0.38$$

approximately; so that the correlation

$$\sqrt{0.38 \times 0.38}$$

instead of being merely 0.38, appeared to be absolute and complete.

Attenuation by errors can also be corrected in another manner, which has the great advantage of an independent empirical basis, and therefore of not being subject to either of the two above mentioned fallacies besetting the other method. Hence, when the results coincide both ways, the fallacies in question may thereby be considered as disproved, for it is very unlikely that they should both be present and in such proportions as to exactly cancel one another. In this method, instead of directly

employing the values $r_{p'p}$, etc., we aggregate them into a single list: by this method, the correlation between some portion of the individual observational errors, and thereby we can see any really existing correspondence to reveal itself in greater completeness. Now, this increase in correlation from this partial elimination of errors will furnish a measure of the increase to be expected from an entire elimination of errors. Assuming the mean error to be inversely proportional both to this increase in the correlation and to the square root of the number of data aggregated, the relation will be:

$$r_{pq} = \frac{\sqrt{r_{p'p} \cdot r_{q'q}}}{\sqrt{n}}$$

connected with q' (or q) quite independently of p or any other link common to p' . Then, the correlation $r_{p'q'}$ will be to that extent increased without any proportional increase in $r_{p'p'}$; hence our above formula will fallaciously present too large a value.

A greater practical difficulty is that of obtaining two series sufficiently independent of one another. For many errors are likely to repeat themselves; even two separate observers are generally, to some extent, warped by the same influences; we are all imposed on by, not only the "Idola Specus," but also the "Idola Tribus" and the "Idola Fori." In such cases, the above formula is still valid, only its correction does not go quite far enough, - a fallacy at any rate on the right side.

An actual instance will best show the urgent necessity of correcting this attenuation. In a correlation between two events, say P and Q , I obtained three independent observations both of P and of Q . The average correlation for those of P with those for Q was 0.38 ($= r_{p'q'}$); the average correlation of those for P with one another was 0.58 ($= r_{p'p'}$); the same for Q was 0.22 ($= r_{q'q'}$). Therefore, the correspondence between the real events, P and Q comes by reckoning to

$$\frac{0.38}{\sqrt{0.58 \times 0.22}} = \text{approximately } 1; \text{ so the the correspondence}$$

instead of being merely 0.38, appeared to be absolute and complete.

Attenuation by errors can also be corrected in another manner, which has the great advantage of an independent empirical basis, and therefore of not being subject to either of the two above mentioned fallacies besetting the other method. Hence, when the results coincide both ways, the fallacies in question may thereby be considered as disproved, for it is very unlikely that they should both be present and in such proportions as to exactly cancel one another. In this method, instead of directly employing the values $p_1 p_2 p_3$, etc. we amalgamate them into a single list; by this means we clearly eliminate some portion of the individual observational errors, and thereby we cause any really existing correspondence to reveal itself in greater completeness. Now, this increase in correlation from this partial elimination of errors will furnish a measure of the increase to be expected from an entire elimination of errors. Assuming the mean error to be inversely proportional both to this increase in the correlation and to the square root of the number of lists amalgamated, the relation will be:

$$r_{pq} = \frac{\sqrt[4]{mn \cdot r_{p'q'}}}{\sqrt[4]{mn}} = 1$$

connected with p (or q) quite independently of p or any other link common to p . Then, the correlation $r_{p'p'}$ will be to that extent increased without any proportional increase in $r_{p'p'}$; hence our above formula will fallaciously present too large a value.

A greater practical difficulty is that of obtaining two series sufficiently independent of one another. For many errors are likely to repeat themselves; even two separate observers are generally, to some extent, warped by the same influences; we are all impressed on by, not only the "Idola Speciei," but also the "Idola Tribus" and the "Idola Voci." In such cases, the above formula is still valid, only its correction does not do quite far enough - a fallacy at any rate on the right side. An actual instance will best show the urgent necessity of correcting this attenuation. In a correlation between two events, say P and Q , I obtained three independent observations both of P and of Q . The average correlation for those of P with those for Q was 0.38 ($= r_{p'p'}$); the average correlation of those for P with one another was 0.58 ($= r_{p'p'}$); the same for Q was 0.52 ($= r_{q'q'}$). Therefore, the correspondence between the real events, P and Q , comes by reckoning to

$$\frac{0.38}{\sqrt{0.58 \times 0.52}} = \text{approximately } 0.52$$

instead of being merely 0.38 , appeared to be absolute and complete.

Attention by errors can also be directed in another manner, which has the great advantage of an independent empirical basis, and therefore of not being subject to either of the two above mentioned fallacies besetting the other method. Hence, when the results coincide both ways, the fallacies in question may thereby be considered as disproved, for it is very unlikely that they should both be present and in such proportions as to exactly cancel one another. In this method, instead of directly employing the values p and q , etc. we amalgamate them into a single list; by this means we clearly eliminate some portion of the individual observational errors, and thereby we cause any really existing correspondence to reveal itself in greater completeness. Now, this increase in correlation from this partial elimination of errors will furnish a measure of the increase to be expected from an entire elimination of errors. Assuming the mean error to be inversely proportional both to this increase in the correlation and to the square root of the number of lists amalgamated, the relation will be:

$$r_{p'p'} = \frac{\sqrt{r_{p'p'}}}{\sqrt{n}}$$

where m and n = the number of independent gradings for p and q respectively,

$r_{p'q'}$ = the mean correlation between the various gradings for p and those for q ,

and $r_{p''q''}$ = the correlation of the amalgamated series for p with the amalgamated series for q .

In the above quoted instance, the three observations for series P were amalgamated into a single list, and similarly those for series Q . Upon this being done, the two amalgamated lists now presented a correlation with one another of no less than $0.66 (= r_{p''q''})$. Thus by this mode of reckoning, the real correspondence became

$$= \frac{\sqrt[4]{3 \times 3 \times 0.66 - 0.38}}{\sqrt[4]{3 \times 3 - 1}} = \text{once more approximately } 1,$$

so that this way also the correspondence advanced from 0.38 to absolute completeness.

If more than two independent series of observations are available, we may acquire additional evidence by trying the effect of partial amalgamation. Instead of throwing all our obtained values together, we may form a set of smaller combinations for each of the two compared characteristics, and then see the mean correlation between one set and the other. In our above instance instead of summarily considering $p'_1 p'_2 p'_3$, we can have $p'_1 p'_2$, $p'_1 p'_3$ and $p'_2 p'_3$ and find out their mean correlation with similar values for q . This works out actually to 0.55 . Hence

$$r_{pq} = \frac{\sqrt[4]{2 \times 2 \times 0.55 - 0.38}}{\sqrt[4]{2 \times 2 - 1}} = \text{approximately } 1.$$

Thus again, by this third way, where both terms are the mean of 9 observed correlational values, the correspondence once more rises from the apparent 0.38 to the real 1 . (1)

3. Limits of Associative Problems.

We have seen that "the length of the arm is said to be correlated with that of the leg, because a person with a long arm has usually a long leg and conversely;" also that this correlation is defined mathematically by any constant which deter-

(1) The exactness of the coincidence between the two methods of correction is in the above instance neither greater nor less than generally occurs in practice. It was singled out, in order to show that the formulae still hold perfectly good even for such an enormous rise as from 0.38 to 1 . The possibility of such a rise is due to the unusual conditions of the experiment in question, whereby the three observations of the same objective series presented the extraordinarily small intercorrelation of 0.22 .

where m and n = the number of independent gradings for p and q respectively.

$r_{p,q}$ = the mean correlation between the various gradings for p and those for q .

and $r_{p,q}''$ = the correlation of the amalgamated series for p with the amalgamated series for q .

In the above quoted instance, the three observations for series p were amalgamated into a single list, and similarly those for series q . Upon this being done, the two amalgamated lists now presented a correlation with one another of no less than 0.88 ($r_{p,q}''$). Thus by this mode of reckoning,

the real correspondence became

$$r_{p,q}'' = \frac{\sqrt{3} \times 0.88 - 0.38}{\sqrt{3} - 1} = \text{one more approximately 1.}$$

so that this way also the correspondence may need from 0.38 to absolute completeness.

If more than two independent series of observations are available, we may acquire additional evidence by trying the effect of partial amalgamation. Instead of throwing all our obtained values together, we may form a set of smaller combinations for each of the two compared characteristics, and then see the mean correlation between one set and the other. In our above instance instead of numerically considering p_1, p_2, p_3 we can have p_1, p_2 and p_3 and find out their mean correlation with similar values for q . This works out actually to 0.35. Hence

$$r_{p,q}'' = \frac{\sqrt{2} \times 0.35 - 0.38}{\sqrt{2} - 1} = \text{approximately 1.}$$

Thus again, by this third way, where both p and q are the mean of 9 observed correlational values, the correspondence once more rises from the apparent 0.38 to the real 1. (1)

3. Limits of Associative Problems.

We have seen that "the length of the list is said to be correlated with that of the leg, because a person with a long arm has usually a long leg and conversely;" also that this correlation is defined mathematically by any constant which deter-

(1) The exactness of the coincidence between the two methods of correlation is in the above instance neither greater nor less than generally occurs in practice. It was singled out, in order to show that the formulae still hold perfectly good even for such an enormous rise as from 0.38 to 1. The possibility of such a rise is due to the unusual conditions of the experiment in question, whereby the three observations of the same objective series presented the extraordinarily small intercorrelation of 0.38.

Mimes the function of any definite size of arm to the mean of the sizes of the corresponding legs. These terms, taken literally, are very wide reaching and express what we will call the "universal" correlation between the two organs.

But evidently not the most painstaking investigation can possibly secure any adequately representative sample for such universal correlations, even in the simple case of arms and legs. To begin with, they would have to be equally derived from every stage of growth, including the prenatal period; since this is the most influential of all causes of variation in size. In the next place, they would have to come from every historical epoch, containing their fair proportion of big Cro-Magnons, little Fufboozers, etc. Further, they must impartially include every living race, from the great Patagonians to the diminutive M'Kabbas; also every social class, from the tall aristocrats to the undermized slummers.

Practically, then, the universal correlation, even if desirable, is quite inaccessible. We are forced to successively introduce a large number of restrictions: the sample is confined to adults, to moderns, to some particular country, etc., etc. In a word, we are obliged to deal with a special correlation.

When we proceed to more narrowly consider these restrictions, it soon becomes clear that they are far from being really detrimental. For every serious investigation will be found to be directed, however vaguely and unconsciously, by some hypothesis as to the causes both of the correspondence and of the digression therefrom. (see page 74). This hypothesis will determine a particular system of restriction, such as to set the correspondence in the most significant relief.

But from these restrictions will at the same time proceed several kinds of grave errors. In the first place, since the restrictions are not explicitly recognized, they often are not carried out in a manner scientifically profitable; ~~they~~ then, the result, however true, may nevertheless be trivial and unsuggestive. For instance, a series of experiments was recently executed by one of our best known psychologists and ended - to his apparent satisfaction - in showing that some children's school-order was largely, correlated with their height, weight and strength. As, however, no step had been taken to exclude the variations due to difference of age, the only reasonable conclusion seemed to be that as children grow older they turned out in fact to probably be the true and sufficient one.

The next fault to be feared is equivocality. For even if the controlling under-thought be good, yet its indistinctness in the mind of the experimenter caused the restriction to be carried out so unsystematically, that the results inevitable ~~may~~ become ambiguous and fruitless.

The last is that, even with the clearest purpose, this specialization of the correlation is an exceedingly difficult matter to execute successfully/ Only by a profound knowledge of the many factors involved, can we at all adequately exclude those irrelevant to our main intention.

Now, all such elements in a correlation as are foreign to the investigator's explicit or implicit purpose will, like the attenuating errors, constitute imputities in it and will quantitatively falsify its apparent amount. This will chiefly happen in two ways.

When the position of the object is changed, the corresponding image is also changed. The image will only be formed when the object is placed at the focal point.

It is also possible to form a virtual image. This occurs when the object is placed between the lens and the focal point. The image is upright and larger than the object. It is also possible to form a real image. This occurs when the object is placed at a distance greater than the focal length. The image is inverted and can be projected onto a screen.

The distance between the object and the lens is called the object distance. The distance between the lens and the image is called the image distance. The focal length of the lens is the distance between the lens and the focal point.

There are two types of lenses: convex and concave. A convex lens is thicker in the middle than at the edges. It can converge light rays. A concave lens is thinner in the middle than at the edges. It can diverge light rays.

The rays of light that pass through a lens are called rays. There are three types of rays: parallel rays, focal rays, and central rays. Parallel rays are rays that are parallel to the principal axis. Focal rays are rays that pass through the focal point. Central rays are rays that pass through the center of the lens.

The rays of light that pass through a lens are called rays. There are three types of rays: parallel rays, focal rays, and central rays. Parallel rays are rays that are parallel to the principal axis. Focal rays are rays that pass through the focal point. Central rays are rays that pass through the center of the lens.

The rays of light that pass through a lens are called rays. There are three types of rays: parallel rays, focal rays, and central rays. Parallel rays are rays that are parallel to the principal axis. Focal rays are rays that pass through the focal point. Central rays are rays that pass through the center of the lens.

The rays of light that pass through a lens are called rays. There are three types of rays: parallel rays, focal rays, and central rays. Parallel rays are rays that are parallel to the principal axis. Focal rays are rays that pass through the focal point. Central rays are rays that pass through the center of the lens.

4. "Constriction" and "Dilation."

Any correlation of either of the considered characteristics will have been admitted irrelevantly, if it has supervened irrespectively of the original definition of the correspondence to be investigated. The variations are thereby illegitimately constrained to follow some irrelevant direction so that (as in the case of Attenuation) they no longer possess full amplitude of possible correlation in the investigated direction; the maximum instead of being 1 will be only a fraction, and all the lesser degrees of correspondence will be similarly affected; such a falsification may be called "constriction." Much more rarely, the converse or "dilation" will occur, by correlations being irrelevantly excluded. The disturbance is measureable by the following relation:

$$r_{pq} = \frac{r_{pq}}{\sqrt{1 - r_{pv}^2}}$$

where r_{pq} = the apparent correlation of p and q, the two variables to be compared,

r_{pv} = the correlation of one of the above variables with a third and irrelevantly admitted variable v.

and r_{pq} = the real correlation between p and q after compensating for the legitimate influence of v.

Should any further irrelevant correlation, say r_{pw} , be admitted, then

$$r_{pq} = \frac{r_{pq}^1}{\sqrt{1 - r_{pv}^2 - r_{pw}^2}}$$

In the reverse case of "dilation,"

$$r_{pq} = r_{pq}^1 \sqrt{1 - r_{pv}^2 - r_{pw}^2}$$

these formulae will be easily seen to be at once derivable from the relations stated on pages 74 and 75. Small, irrelevant variations evidently do not affect the result in any sensible degree, while large ones are capable of revolutionizing it.

The following is an actual illustration of this constriction. I was investigating the correspondence between on the one hand intelligence of school lessons and on the other the faculty of discrimination of musical pitch. The correlation proved to be 0.49. But, upon inquiry, it turned out that more than half of the children took lessons in music and therefore enjoyed artificial training as regards pitch; here, then, was a powerful cause of variation additional and quite irrelevant to the research, which dealt with the correspondence between the two natural faculties. When this disturbance had once been detected, there was no difficulty in eliminating its influence by the above formula; the correspondence between pitch and discrimination and music lessons was measured at 0.61; so that the true required correlation became

0.49

+ 0.62.

$$\frac{0.49}{\sqrt{1 - 0.61^2}}$$

4. "Constitutional" and "Political" .

[illegible]

 

[illegible][illegible]

In this particular case, the more desirable course was open of elimination of the constricting, practically, by confining the experiment to those children who were learning music and therefore were on a sufficient equality as regards the training. The correlation then gained in this purely empirical way exactly coincided with the former results, being again 0.62.

5. "Distortion."

Whereas Attention and Constriction have wholly tended to reduce the apparent correlation, and Dilation to enlarge it, we now come to a third kind of impurity that may equally well reduce or enlarge. Its effects is thus analogous to the first consequence of accidental errors discussed in the first part of this article, but unlike the latter, this Distortion does not in the least tend to eliminate itself in the longest series of observations.

Distortion occurs whenever the two series ~~are~~ to be compared together both correspond to any appreciable degree with the same third irrelevant variant. In this case, the relation is given by

$$r_{pq}^1 = \frac{r_{pq}^1 - r_{pv} \cdot r_{qv}}{\sqrt{(1 - r_{pv}^2)(1 - r_{qv}^2)}} \quad (1)$$

where r_{pq}^1 = the apparent correlation between p and q, the two characteristics to be compared,

r_{pv} and r_{qv} = the correlations of p and q with some third and perturbing variable v.

and r_{pq} = the required real correlation between p and q after compensating for the illegitimate influence of v.

Should the common correspondence with v have been irrelevantly excluded instead of admitted, the relation becomes

$$r_{pq} = r_{pq}^1 \sqrt{(1 - r_{pv}^2)(1 - r_{qv}^2)} + r_{pv} \cdot r_{qv}$$

In the course of the same investigation above alluded to, but in another school, the correlation between school intelligence and discrimination of pitch turned out to be - 0.25, so that apparently not the cleverer but the stupider children could discriminate best! But now it was observed that a superiority in discrimination had been shown by the older children, amounting to a correlation of 0.55; while, for a then unknown reason, the schoolmaster's estimate of intelligence had shown a very marked (though unconscious) partiality for the younger ones, amounting to a correlation of 0.65. Hence the true correlation reckoned out to

$$\frac{-0.25 - 0.55 \times (-0.65)}{\sqrt{(1 - 0.55^2)(1 - 0.65^2)}}$$

= + 0.17. This latter low but direct correlation was - under the particular

1. This same formula has already been arrived at, though though along a very different route, by Yule. See Proc. R.S.L. Vol. IX.

circumstances of the experiment - unquestionably about correct; so that the one originally observed of $= -0.25$ would have been entirely misleading.

6. Criticism of Prevalent Working Methods.

So far, our illustration of systematic deviation has been confined to instances taken from personal experience. But it might perhaps be thought that other workers avoid such perversions of fact by the simpler method of common sense. Unfortunately, such does not seem to have been at all the case; not once, to the best of my knowledge, has any partial association between two psychological events been determined in such a way as to present any good evidential value - these are strong terms, but I think, hardly exaggerated.

Psychologists, with scarcely an exception, never seem to have become acquainted with the brilliant work being carried on since 1886 by the Galton-Pearson school. The consequence has been that they do not even attain to the first fundamental requisite of correlation, namely, a precise quantitative expression. Many have, indeed, taken great pains in the matter and have constructed arrays of complicated numerical tables; but when we succeed in orienting ourselves in the somewhat bewildering assemblage of figures, we generally find that they have omitted precisely the few facts which are essential, so that we cannot even work out the correlation for ourselves.

This lack of quantitative expression entails far more than merely diminished exactitude. For in consequence, the experimenters have been unable to estimate their own results at all correctly, some have believed themselves to demonstrate an entire absence of correspondence, when the latter has really been quite considerable; whereas others have presented to the public as a high correlation what has really been very small and often well within the limits of mere accidental coincidence; these limits they have had no means of determining, and moreover their data were usually obtained in such a way as to make it unnecessarily large.

Seeing, thus, that even the elementary requirements of good correlation work described in the first part of this article have been so generally deficient we cannot be surprised to find that the more advanced refinements of procedure discussed in the second part have been almost wholly unregarded; so that the final results are saturated and falsified with every description of impurity. In this respect, unfortunately, it is no longer possible to hold up even the Galton-Pearson school as a model to be imitated. The latter must now perform the very different office of saving us from detailed criticism of inferior work, by enabling us to form an opinion as to how far the defect permeates and vitiates even the best existent correlational research.

As example, we will take Pearson's chief line of investigation, Collateral Heredity, at that point where it comes into closest contact with our own topic, Psychology. Since 1889 he has, with government sanction and assistance, been collecting a vast number of data as to the amount of correspondence existing between brothers. A preliminary calculation, based in each case upon 800 to 1,000 pairs, led, in 1910 to the publication of the following momentous results:

... ..
... ..

[illegible][illegible]

Coefficients of Collateral Heredity.
Correlation of Pairs of
Brothers.

<u>Physical Characters.</u>		<u>Mental Characters.</u>	
(Family Measurements)		(School Observations)	
Stature	0.5107	Intelligence	0.4559
Forearm	0.4912	Vivacity	0.4702
Span	0.5494	Conscientiousness	0.5929
Eye-colour	0.5169	Popularity	0.5044
		Temper	0.5068
		Self-consciousness	0.5915
		Shyness	0.5281
(School Observations)			
Cephalic index	0.4861		
Hair Colour	0.5452		
Health	0.5303		
Mean	0.5171	Mean	0.5214

Dealing with the means for physical and mental characters, we are forced to the perfectly definite conclusion that the mental characters in men are inherited in precisely the same manner as the physical. Our mental and moral nature is, quite as much as our physical nature, the outcome of heredity factors.

Now let us consider how these coefficients of correlation will be affected by our "systematic deviations." To begin with there is the "Attenuation" by errors; since it evidently cannot be assumed that the schoolmasters' judgments as to conscientiousness, temper, etc. are absolutely infallible. On page 90, it has been shown that deviation from this source may be estimated by the following formula:

$$r_{pq} = \frac{r_{p'q'}}{\sqrt{r_{p'p'} \cdot r_{q'q'}}$$

To ascertain $r_{p'p'}$ and $r_{q'q'}$, I am aware of no precise data ~~between~~ beyond that found in some experiments of my own, where the independent intellectual gradings for the same series of subjects correlated with one another on an average to the amount of -.64. As on other occasions very competent persons have estimated this to be as much as should be expected, as intelligence is about the most easily gradable of all the mental qualities mentioned by Pearson, there is so far no reason to suppose that his "great number of masters and mistresses" did on the whole any better. Hence, even if we could assume that the mistakes in estimating one brother were independent of the mistakes in estimating the other, then the true correlation would be about, not 0.5172,

but $\frac{0.5172}{\sqrt{0.64 \times 0.64}} = 0.81$, an extent of difference that seriously

modifies our impression of exactitude from all these coefficients to four places of decimals. When we further consider that each of these physical and

UNITED STATES DEPARTMENT OF AGRICULTURE
OFFICE OF THE SECRETARY
WASHINGTON, D. C.

Statement of

(Name of Person)

Witnessed and Subscribed

(Name of Person)

1917.
 1918.
 1919.
 1920.
 1921.
 1922.
 1923.
 1924.

1917.
 1918.
 1919.
 1920.
 1921.
 1922.
 1923.
 1924.

1917.

1917.

I, the undersigned, do hereby certify that the foregoing is a true and correct copy of the original as the same appears in the files of the Department of Agriculture.

Witness my hand and the seal of the Department of Agriculture at Washington, D. C., this 1st day of January, 1924.

Secretary

 1924

I, the undersigned, do hereby certify that the foregoing is a true and correct copy of the original as the same appears in the files of the Department of Agriculture.

Witness my hand and the seal of the Department of Agriculture at Washington, D. C., this 1st day of January, 1924.

I, the undersigned, do hereby certify that the foregoing is a true and correct copy of the original as the same appears in the files of the Department of Agriculture.

Witness my hand and the seal of the Department of Agriculture at Washington, D. C., this 1st day of January, 1924.

mental heredity can hardly be more than mere accidental coincidence.

Let us next proceed to irrelevant correlation, and take for our theme postnatal accidents connected on the one side with brotherhood and on the other with the mental qualities. Pearson's primary intention seems to have been to make his correlations as "universal" as possible and in one place he expressly mentions that education is among the causes contributory to variation. Hence, he is more than consistent, in that he forms his correlation without regard to the fact that the ~~correlation~~ correspondence between the brothers' "conscientiousness," "popularity," etc., must be in great measure due to their coming under the same home influences. But such a correlation can scarcely be accepted as scientifically valuable. For we do not really know anything precise about the assimilating effects of heredity, when our observed correspondence is perhaps chiefly due to the brothers having the same amount of hampers and pocket-money. Still less can we, then, fairly compare such results with that obtained from physical measurements, where common home life has little or no effect. The factor of post-natal accidents, therefore, cannot but be regarded as irrelevant, and consequently the coefficients of correlation must be taken as hopelessly "distorted."

But even consistence cannot be upheld throughout the matter. For though the effect of postnatal life has thus been admitted with regard to education at home, it has perforce been excluded as regards public education. For only those brothers have been compared together who are at the same school; the coefficients of correlation would certainly diminish if those also could be included who are living in a totally different manner, have gone to sea, etc., The correlations are therefore also illegitimately "dilated."

If this work of Pearson has thus been singled out for criticism, it is certainly from no desire to undervalue it. The above and any other systematic errors are eventually capable of adequate elimination, and this article has itself, it is hoped been of some use towards that purpose. Such correction will no doubt necessitate an immense amount of further investigation and labour, but in the end his results will acquire all their proper validity. My present object is only to guard against premature conclusions and to point out the urgent need as still further improving the existing methodics of correlation work, a method of investigation which he himself has so largely helped to create and by means of which he is carrying light into immense regions hitherto buried in the obscurity of irresponsible speculation. The fundamental difference between his procedure and that here recommended, is that he seeks large natural samples of any existing series sufficiently homogeneous to be treated mathematically; whereas here smaller samples are deemed sufficient, but they are required to be artificially selected, ordered, and corrected into full scientific significance. His methods are those of pure statistics these inculcated here may be more aptly termed "statisticoids."

7. Number of Cases Desirable for an Experiment.

This leads us to the important question, as to how many cases it is advisable to collect for a single series of experiments. In actual practice, the greatest diversity has been apparent in this respect; many have thought to sufficiently establish important correlations with less than ten experimental subjects while others have thought it necessary to gather together at least over a thousand.

of correlation between the two variables. The correlation coefficient is a measure of the strength and direction of the relationship between two variables. It ranges from -1 to +1. A value of +1 indicates a perfect positive correlation, a value of -1 indicates a perfect negative correlation, and a value of 0 indicates no correlation. The correlation coefficient is calculated using the following formula:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

where r is the correlation coefficient, x_i and y_i are the individual data points, $\bar{x}$ and $\bar{y}$ are the means of the two variables, and $\sum$ is the summation symbol. The correlation coefficient is a useful tool for analyzing the relationship between two variables, but it is important to remember that it only measures the linear relationship between the two variables. It does not measure the strength of the relationship in general, and it does not indicate causation.

There is a great deal of talk about the "new" and "old" world, but the fact is that the world is not new or old, it is just different. The world is a big place, and it is full of people who are different from each other. But that is what makes it so interesting. We should not be afraid of differences, we should embrace them. We should learn from each other and grow together. The world is our home, and it is our responsibility to make it a better place for everyone.

[illegible][illegible]

Now, a series of experiments is a very limited extract, whose disposition is, nevertheless, to be accepted as a fair sample of the whole immense remainder. Other things equal, then, the larger the sample the greater its evidential value and the less chance of a mere occasional coincidence being mistaken for the permanent universal tendency.

This danger of accidental deviation has been discussed in the first part and there shown to be strictly measurable by the "probable error." We there saw, also, that this danger can never be entirely eliminated by any sample however large, so that it is necessary to accept some standard less rigorous than absolute certainty as sufficient for all practical purposes; usually, the danger of mere chance coincidence is considered to be inappreciable when a correlation is observed as much as five times greater than the probable error, seeing that mere chance would not produce this once in a thousand times. Hence, evidently, the accidental deviation depends, not only on the number of cases, but also on the largeness of the really existing correspondence; the more perfect the latter, the fewer the cases that will be required to demonstrate it conclusively; and this tendency is augmented by the fact that the probable error, besides varying inversely with " n ", does so to a further extent with " n^2 " (see formula). It was shown in the same part that the size of the probable error also varies accordingly to the method of calculation- and to such an extent that twenty cases treated in one of the ways described furnishes as much certitude as 180 in another more usual way. If the common tri-fold classification be adopted, an even greater number is required to effect the same purpose; and if the correlation be not calculated quantitatively at all, but instead be presented in the customary fashion to the reader's general impression, then no number of cases whatever appear sufficient to give reasonable guarantee of proof.

While this the number of subjects is not by any means the sole factor in diminishing even the accidental deviation it has no effect whatever upon the far more formidable systematic deviation except that it indirectly leads to an enormous augmentation thereof. When we are taking great pains to be able to show upon paper an imposing number of cases and a diminutive probable error, we are in the self same process most likely introducing a systematic deviation twenty times greater.

From this we may gather that the number of cases should be determined by the simple principle, that the measurements to be aggregated together should have their error brought to the same general order of magnitude. An astronomical chronometer, with spring-detent escapement, is not the best travelling clock; nor is there any real advantage in gracing upon a milestone (as has actually been done by an infatuated mathematician!) the distance to the nearest village in metres to three decimal places. Now, the present stage of Correlational Psychology is one of pioneering; and, instead of a few unyielding experiments, we require a large number of small ones carefully carried out under varied and well considered conditions. At the same time, however, the probable error must be kept down to limits at any rate small enough for the particular object of investigation to be proved. For such a purpose a probable error may at present be admitted without much hesitation up to about 0.05; so that, by adopting the method of calculation recommended, two to three dozen subjects should be sufficient for most purposes. The precision can always be augmented subsequently, by carrying our similar experiments under similar conditions and then taking averages. Only after a long preliminary exploration of this rougher sort, shall we be in a position to effectually utilize experiments designed and executed from the very beginning on a vast scale.

PLEASE DO NOT REMOVE
CARDS OR SLIPS FROM THIS POCKET

UNIVERSITY OF TORONTO LIBRARY

BF Spearman, Charles Edward
181 The proof and measurement of
S73 association between two things